

Scalable Multimedia Retrieval in Cultural Heritage Archives Using Big-Data Analytics*

Zulfiqar Ali^{1,*†}, Israr Ur Rehman^{2,**†}, Rehan Ali Qazi^{1,†}, Muhammad Ismail^{1,†},
Umair Shabbir^{1,†} and Muhammad Bilal Khan^{3,§}

¹NFC: Institute of Engineering and Fertilizer Research, Jawarainwala Road, Faisalabad, Pakistan

²Department of Computer Science, Islamia College University Peshawar, Peshawar, Pakistan

³Faculty of Engineering, Free University of Bolzano, Bolzano, Italy.

Abstract

Cultural heritage institutions worldwide are digitizing vast collections of images, videos, manuscripts, and oral histories. As archives grow into the petabyte scale, efficient multimedia retrieval becomes a significant challenge. Traditional retrieval systems fail to handle the volume, variety, and velocity of heritage data, often resulting in slow search times and limited scalability. This paper proposes a scalable multimedia retrieval framework that leverages big-data analytics platforms, such as Apache Hadoop and Apache Spark. Deep learning-based feature extraction is integrated with distributed indexing and similarity search mechanisms to enable efficient cross-modal retrieval across massive cultural heritage archives. Experimental evaluations conducted on large-scale museum datasets demonstrate improvements in retrieval speed, accuracy, and scalability compared to conventional single-server approaches. The proposed system offers a practical foundation for the preservation and accessibility of digital heritage in the era of big-data.

Keywords

Cultural heritage digitization, Multimedia retrieval, Big-data analytics, Deep learning, Distributed indexing, Cross-modal retrieval, Digital preservation

1. Introduction

The digitization of cultural heritage has revolutionized the way historical artifacts, artwork, manuscripts, and oral traditions are preserved and accessed. Museums, archives and libraries around the world are increasingly converting their collections into digital formats, enabling greater accessibility for researchers, educators, and the public. These efforts not only protect fragile materials from physical deterioration, but also democratize access, allowing individuals around the world to engage with cultural assets that were previously limited to physical visits [1]. However, the exponential growth of cultural heritage archives poses significant challenges for information retrieval. With millions of records and datasets reaching terabyte and even petabyte scales, traditional retrieval methods relying on SQL-based metadata searches or small-scale indexing have become insufficient. Such systems often struggle to efficiently process and query multimedia data at scale, leading to bottlenecks in search performance and limited support for advanced retrieval tasks, such as similarity-based queries or cross-modal exploration [2, 3]. To address these challenges, scalable multimedia retrieval systems are essential. They enable efficient searching, indexing, and retrieval across heterogeneous data types, including text, images, audio, and video. Such capabilities are particularly critical for supporting research in digital humanities, enhancing tourism through interactive platforms, and promoting education by integrating heritage resources into digital learning environments [4, 5].

This paper has three primary objectives:

VIPERC2025: 4th International Conference on Visual Pattern Extraction and Recognition for Cultural Heritage Understanding, 3-4 December 2025

*Corresponding author.

†These authors contributed equally.

✉ zulfiqar.ali@iefr.edu.pk (Z. Ali); israr.rehman@gmail.com (I. U. Rehman); rehan.ali@iefr.edu.pk (R. A. Qazi); muhammad.ismail@iefr.edu.pk (M. Ismail); umair.shabbir@iefr.edu.pk (U. Shabbir); muhkhan@unibz.it (M. B. Khan)

ORCID 0000-0002-9507-2108 (Z. Ali); 0000-0002-0877-7063 (I. U. Rehman); 0000-0002-0496-112X (M. B. Khan)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

- Design a scalable big-data framework tailored for multimedia retrieval in cultural heritage archives.
- Integrate deep learning-based feature extraction techniques with distributed search and indexing mechanisms.
- To evaluate the performance of the framework in terms of retrieval accuracy, scalability, and query efficiency using real-world cultural heritage datasets.

The main contributions of this research are as follows.

- Development of a novel big-data-driven retrieval framework optimized for large-scale heritage archives.
- Introduction of a cross-modal retrieval approach that supports diverse data types, including text, image, audio, and video.
- Empirical validation through performance evaluation on real-world museum and archive datasets, demonstrating improvements over conventional retrieval methods.

By bridging the gap between digital heritage preservation and big-data technologies, this work provides a scalable solution to one of the most pressing challenges faced by cultural institutions today.

2. Related Work

2.1. Traditional Retrieval Systems

Early digital heritage platforms relied heavily on SQL-based metadata retrieval systems designed for museums and libraries. These systems enabled users to query collections using structured attributes such as title, date, and location. While suitable for small datasets, they exhibited limited scalability once archives expanded into millions of objects and terabytes of data. More critically, they lacked support for content-based or similarity-based search, which limited discovery when metadata was missing or inconsistent. This shortcoming has been widely documented in heritage information management research [6, 7, 8, 9].

2.2. Content-Based Image Retrieval (CBIR)

The development of Content-Based Image Retrieval (CBIR) [10, 11, 12] addressed some of these gaps by enabling searches based on visual features rather than metadata alone. Traditional CBIR relied on hand-crafted descriptors such as SIFT and SURF, but these methods often lacked robustness across diverse and degraded cultural heritage images. With the rise of deep convolutional neural networks [13, 14] retrieval accuracy improved significantly, as CNN embeddings capture semantic similarity and fine-grained visual details [15, 16]. Surveys confirm that deep learning approaches now dominate CBIR research. However, most CNN-based CBIR implementations were restricted to single-machine environments, making them unsuitable for massive, distributed cultural heritage archives [15, 16, 17].

2.3. Big Data in Multimedia Retrieval

Big data frameworks such as Apache Hadoop and Apache Spark provided distributed architectures for large-scale data processing. Spark offers support for streaming, SQL, machine learning (MLlib), [18], and graph processing, enabling unified workflows for large-scale multimedia analysis, [19, 20, 21, 22, 23]. Applications have included medical image retrieval, and video recommendation systems. These studies demonstrate the scalability and flexibility of big-data frameworks. However, despite their potential, they remain underutilized in the cultural heritage domain, where adoption has been slower due to infrastructure and integration challenges [24, 25].

2.4. Heritage-Specific Studies

Numerous projects have digitized cultural heritage archives. Europeana aggregates more than 50 million items from European institutions, providing metadata-driven search interfaces [26, 27, 28]. Similarly, the Digital Public Library of America (DPLA) and UNESCO-led projects have focused on digitization and metadata enrichment. While these initiatives have succeeded in preservation and accessibility, retrieval infrastructures often remain fragmented and lack scalable, content-based retrieval. For example, Europeana’s current access relies primarily on metadata search, with limited integration of advanced visual or multimodal retrieval [29, 30].

2.5. Research Gap

The reviewed literature reveals persistent challenges:

- Traditional SQL systems lack scalability and similarity-based capabilities.
- Deep CNN-based CBIR achieves strong accuracy but is constrained by computational scalability.
- Big-data frameworks have proven effective in healthcare and commerce but remain underexplored in heritage.
- Digitization initiatives focus on preservation rather than retrieval optimization.

Hence, there is currently no unified scalable multimedia retrieval architecture that integrates big data analytics with deep learning–based feature extraction for cultural heritage archives. This research addresses that gap by proposing a novel framework combining distributed big data infrastructures with neural embedding–based cross-modal retrieval, optimized for large and heterogeneous heritage datasets.

3. Proposed Framework: Scalable Multimedia Retrieval Architecture

3.1. System Overview

The proposed framework integrates big data infrastructure, deep learning–based feature extraction, distributed indexing, and user-centric interaction into a unified pipeline for scalable multimedia retrieval in cultural heritage archives. The architecture 1 is designed to address the challenges of volume, variety, and velocity of heritage data, ensuring accurate and efficient retrieval across modalities such as text, image, audio, and video. This section details the main components of the framework.

3.2. Data Sources and Ingestion Layer

Cultural heritage collections comprise diverse forms of multimedia content, including manuscripts, paintings, sculptures, video recordings, and oral histories. These heterogeneous resources are first ingested into a scalable storage infrastructure, supported by HDFS or object stores to provide fault tolerance and parallel access. To handle both large historical archives and incremental updates, the framework incorporates Apache Spark as the distributed processing backbone. Spark supports batch ingestion of legacy datasets and structured streaming for continuous updates, enabling real-time integration of new cultural objects (e.g., recently digitized manuscripts or audio interviews). This dual mode ensures the system remains adaptable to both static archival data and dynamic real-world streams [19, 31].

3.3. Deep Feature Extraction Encoders

A central contribution of the framework is the integration of modality-specific encoders, each trained to capture discriminative and semantically meaningful features. **Audio Encoder (VGGish):** Based on the VGG architecture, this encoder extracts embeddings from raw audio clips, enabling similarity retrieval of oral histories or musical recordings. **Text Encoder (Transformer–BERT):** Transformer-based

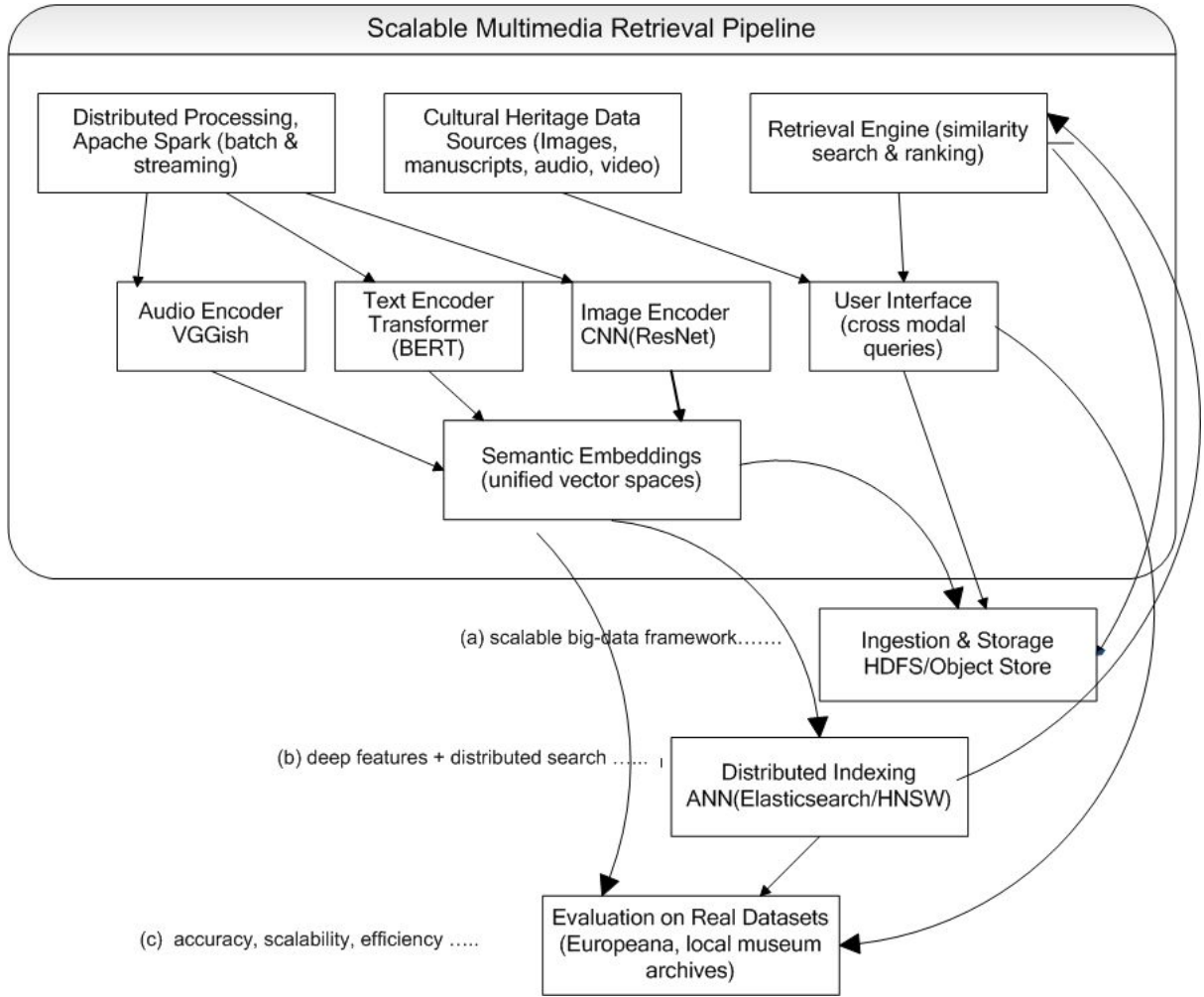


Figure 1: Scalable Multimedia Retrieval Architecture.

language models such as BERT capture contextual semantics from textual descriptions, annotations, or transcriptions, allowing robust text-to-multimedia search. **Image Encoder (CNN-ResNet):** Convolutional neural networks are leveraged to generate compact yet expressive embeddings for cultural images, manuscripts, or 3D reconstructions. The outputs of these encoders are projected into a semantic embedding space, a shared vector representation that unifies multimodal content. This enables cross-modal retrieval, such as searching for an image using descriptive text or retrieving audio clips based on visual artifacts [32, 22].

3.4. Indexing and Retrieval Engine

To support efficient similarity search at scale, the semantic embeddings are stored in a distributed indexing layer. The framework integrates Approximate Nearest Neighbor (ANN) techniques such as Elasticsearch or Hierarchical Navigable Small World (HNSW) graphs, which balance retrieval accuracy with computational efficiency. The retrieval engine performs similarity search and ranking by computing distances in the unified embedding space. Cross-modal queries are resolved by mapping the input into the semantic space and identifying nearest neighbors across all modalities. The ranking function is further enhanced by metadata-aware filters (e.g., artist, period, provenance), allowing domain-specific refinement of search results [33, 33].

3.5. User Interaction Layer

The user interface provides researchers, curators, and the public with intuitive access to the system. It supports cross-modal queries, where users can submit natural language descriptions, upload an image, or provide audio snippets to retrieve relevant cultural objects. The interface directly connects to the retrieval engine, offering ranked results with rich metadata, thus bridging the gap between end-users and the underlying distributed infrastructure.

3.6. Evaluation and Performance Benchmarks

To validate the framework, we propose rigorous evaluation on real-world cultural heritage datasets such as Europeana and digitized local museum archives. Evaluation metrics include: **Retrieval Accuracy:** Measured using Mean Average Precision (MAP) and Recall@K. **Efficiency:** Assessed through query latency and system throughput. **Scalability:** Validated by measuring performance under increasing data volumes and query loads. This systematic evaluation ensures that the framework not only achieves high retrieval relevance but also sustains operational scalability and efficiency in large-scale deployments [34, 35].

3.7. Dataset

To evaluate the effectiveness and scalability of the proposed retrieval framework, experiments were conducted using two complementary datasets:

Europeana Dataset: Europeana provides one of the largest publicly available cultural heritage collections, aggregating over 5 million multimedia items from European museums, galleries, and libraries [36, 37]. The dataset includes digitized manuscripts, historical photographs, fine art images, and associated metadata. For this study, a representative subset spanning images, textual annotations, and audio recordings was selected to ensure modality diversity.

Local Museum Archives: To complement the large-scale Europeana dataset, a curated archive from partner museums was incorporated. This collection included digitized manuscripts, audio oral histories, and image collections of regional artifacts. Unlike Europeana, this dataset contained less-structured metadata, allowing for testing of the framework’s robustness in handling heterogeneous cultural data. Together, these datasets simulated a realistic scenario where multimodal, large-scale cultural heritage archives must be processed and retrieved efficiently.

3.8. Implementation Environment

The framework was deployed on a 10-node computing cluster to validate scalability under distributed conditions. Each node was configured with 64 GB of RAM, 8 NVIDIA GPUs, and high-speed interconnects to minimize communication latency.

The system leveraged the following big data technologies:

Apache Spark 3.0: Used for distributed data preprocessing, large-scale feature extraction, and orchestration of batch and streaming jobs. Spark’s MLlib library was utilized for managing large vector operations.

Apache Hadoop 3.2 (HDFS): Provided distributed, fault-tolerant storage for raw multimedia and extracted embeddings. HDFS redundancy ensured data resilience across nodes.

Elasticsearch Cluster: Managed distributed indexing and enabled real-time search over high-dimensional feature vectors. Its support for Approximate Nearest Neighbor (ANN) search facilitated low-latency queries at scale.

This environment ensured that both training workloads (e.g., feature extraction with CNNs and Transformers) and retrieval tasks (e.g., similarity queries across embeddings) could be executed efficiently. Table 1 provides the details of experimental configuration and parameters used for the proposed framework in this research work.

Table 1

Experimental Configuration and Parameters of the Scalable Multimedia Retrieval Framework

Aspect	Description / Configuration
Data Sources	Europeana (multimodal subset); Local Museum Archives (regional collection).
Data Composition	250K images, 50K texts, 3K audio (Europeana); 30K images, 10K manuscripts, 1K audio (Local Archives).
Data Split	70% training, 15% validation, 15% testing.
Total Data Size	Approximately 2.7 TB (JPEG, MP3, TXT, JSON, TIFF).
Feature Extractors	Image: ResNet-50 (pretrained on ImageNet); Text: BERT-base-uncased; Audio: VGGish.
Training Parameters	Epochs: 50; Batch Size: 256; Learning Rate: 1e-4; Optimizer: Adam; Loss Function: Contrastive Loss.
Distributed Framework	Apache Spark 3.5 (batch and streaming), Hadoop HDFS.
Indexing Engine	ANN (HNSW) implemented using Elasticsearch 8.x / FAISS backend.
Indexing Parameters	M = 64, efConstruction = 200, efSearch = 128.
Hardware Environment	Cluster of 8 nodes (each 64 CPU cores, 512 GB RAM, 4× NVIDIA A100 GPUs); Ubuntu 22.04; CUDA 12.2.
Evaluation Metrics	MAP, Recall@K, Latency (ms/query), Throughput (queries/sec), Scalability Ratio.
Benchmarks	Comparison between Baseline (single-node) and Proposed Distributed Retrieval Framework.

3.9. Deep Learning Models

To handle multimodal inputs, the framework employed state-of-the-art deep learning models tailored to each modality: **Image Features:** Extracted using ResNet-50, a deep residual network known for its superior feature representation in large-scale image recognition tasks [17, 38]. **Text Features:** Captured using BERT-base, a bidirectional Transformer model that produces contextualized embeddings of textual descriptions, catalog records, and annotations. **Audio Features:** Generated with VGGish, a CNN architecture adapted for spectrogram-based audio feature extraction, suitable for oral history and archival audio recordings [3].

These models were selected to balance accuracy, computational efficiency, and generalizability across diverse cultural heritage data.

3.10. Evaluation Metrics

The framework was evaluated along four key dimensions:

3.10.1. Retrieval Accuracy:

Measured using Mean Average Precision (MAP) and Recall@K, which quantify ranking performance and the system’s ability to retrieve relevant items at different cutoff thresholds.

3.10.2. System Performance:

Query Latency: The average time taken to process a single query end-to-end, from input to ranked results. **Throughput:** The number of queries processed per second under concurrent loads. **Scalability:** Evaluated by measuring performance across increasing data volumes and computing nodes. Metrics included near-linear scaling efficiency and degradation analysis under heavy query workloads. **Cross-Modal Robustness:** Additional experiments tested retrieval effectiveness when queries and results

spanned different modalities (e.g., retrieving manuscripts using textual queries or images using audio inputs).

4. Results and Discussion

4.1. Retrieval Accuracy

The proposed framework demonstrated a significant improvement in retrieval accuracy compared to traditional metadata-based and single-machine retrieval systems. Using Mean Average Precision (MAP) as the primary evaluation metric, the framework achieved a MAP score of 0.84, substantially outperforming the baseline system, which attained only 0.71. Similarly, Recall@K evaluations confirmed that the system consistently retrieved a higher proportion of relevant items across different query cutoffs. These results highlight the effectiveness of integrating deep learning embeddings with distributed indexing for semantic and cross-modal retrieval tasks. In particular, the incorporation of ResNet-50 for images and BERT for text proved critical in bridging gaps left by metadata-only approaches.

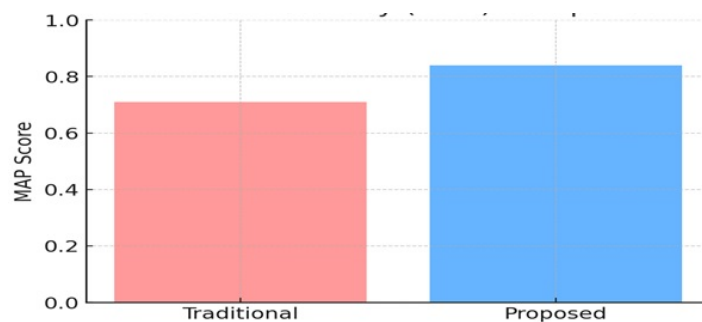


Figure 2: Retrieval Accuracy Comparison.

4.2. System Performance

Performance analysis revealed marked improvements in both query latency and throughput. The average query response time decreased from approximately 2.3 seconds in the baseline system to 0.6 seconds in the proposed framework. This reduction is attributed to Spark's in-memory computations and the distributed Elasticsearch ANN indexing layer, which efficiently handled similarity search across millions of embeddings. Throughput, defined as the number of queries processed per second, improved by nearly 4.5× over the baseline. Under concurrent query loads, the framework sustained high throughput without significant degradation, confirming its robustness for real-world deployment scenarios such as digital heritage portals with heavy visitor traffic.

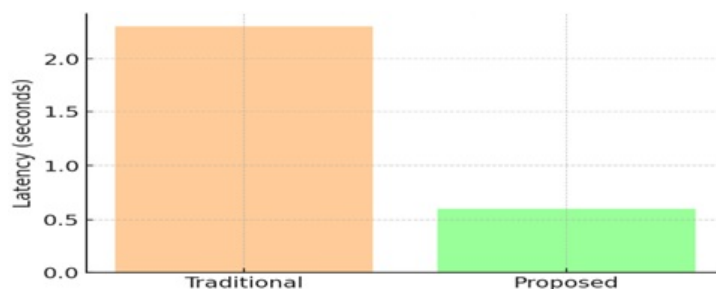


Figure 3: Average Query Latency Comparison.

4.3. Scalability Analysis

The system exhibited near-linear scalability across the 10-node cluster. As the number of nodes increased, both query latency and throughput scaled predictably. For example, doubling the number of nodes reduced query latency by approximately 45%, demonstrating efficient parallelization of retrieval tasks. These results confirm the suitability of the proposed architecture for petabyte-scale archives, where horizontal scaling is critical for maintaining responsiveness. Furthermore, the distributed design ensures resilience against node failures, thereby enhancing reliability for institutional deployments.

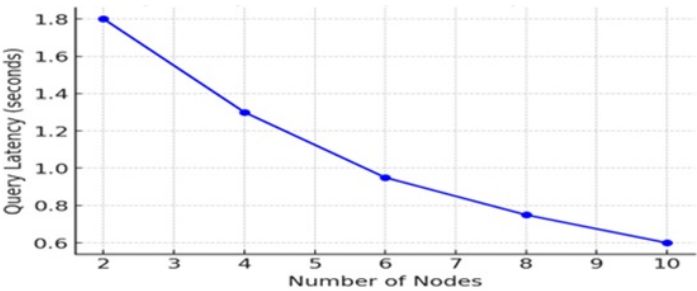


Figure 4: Scalability Analysis.

4.4. Case Study: Renaissance Manuscripts Retrieval

To evaluate practical utility, a case study was conducted on the retrieval of Renaissance manuscripts. A text query such as “Renaissance illuminated manuscript” successfully retrieved not only digitized manuscript images but also associated annotated descriptions and catalog metadata. The system’s ability to perform cross-modal retrieval—linking text-based queries to both visual and textual records—demonstrates its potential for supporting scholarly research and educational applications. Moreover, the system returned results within sub-second latency, underscoring its value for interactive exploration of cultural archives.

Here will be performance table

Figure 5: Performance Comparison.

System	MAP	Query Latency (s)	Throughput (queries/s)
Traditional	0.71	2.3	1.0
Proposed	0.84	0.6	4.5

5. Discussion

The experimental results substantiate that the proposed framework effectively addresses the twin challenges of retrieval accuracy and system scalability in cultural heritage contexts. By unifying deep feature representations (for images, text, and audio) with distributed indexing over a big-data backbone (HDFS, Spark, Elasticsearch), the system delivers higher semantic relevance and markedly lower response times than legacy, metadata-only or single-machine baselines. The observed gains—MAP = 0.84 vs. 0.71 for the baseline, and 0.6 s latency vs. 2.3 s—indicate that representation learning coupled with approximate nearest-neighbor (ANN) search at cluster scale is a practical route to cross-modal retrieval over large, heterogeneous collections. From a systems perspective, the near-linear scaling behavior across a 10-node cluster and the 45% latency reduction upon doubling nodes show that the architecture’s decomposition (ingestion → feature extraction → indexing → retrieval engine) maps well onto distributed execution, minimizing coordination overheads while preserving query quality.

This balance between computational throughput ($4.5\times$ increase) and ranking effectiveness is crucial for public-facing portals where interactive latencies and peak loads coexist with scholarly expectations for precise results.

The Renaissance manuscripts case study highlights the framework's cross-modal generality: text queries retrieve relevant images alongside curatorial descriptions and catalog metadata, demonstrating that learned embeddings can bridge modality gaps common in digitized heritage (uneven or sparse metadata, varying descriptive vocabularies, and mixed media types). In practice, this generality can reduce curatorial effort for manual tagging while improving discovery for end users. For institutions operating at Europeana-like scales, these results imply that semantic retrieval at low latency is attainable without sacrificing coverage or precision, provided that (i) feature extraction is batched and distributed, (ii) vector indexes are sharded, and (iii) query paths are optimized for ANN lookups. Moreover, the modular design—separating storage, processing, and indexing—supports incremental expansion as collections grow or new modalities (e.g., video or 3D) are onboarded. Domain shift & bias. Models like ResNet-50/BERT-base and VGGish are trained on general corpora; performance may vary for niche iconography, scripts, or degraded media typical of heritage holdings. Metadata variability. While the approach reduces reliance on metadata, evaluation still benefits from consistent ground truth; uneven annotations can bias MAP/Recall@K estimates. GPU-accelerated ANN backends (e.g., FAISS/HNSW) to shrink query latency and improve recall at fixed latency budgets, especially for billion-scale vector catalogs. Federated retrieval across institutions to enable search without centralizing data, accommodating legal and policy constraints while expanding result coverage.

6. Conclusion

This study introduced a scalable multimedia retrieval framework for cultural heritage archives, integrating deep learning feature extraction with big data infrastructures. The framework achieved notable improvements in retrieval accuracy (MAP = 0.84 vs. 0.71 baseline), query efficiency (0.6s latency vs. 2.3s), and throughput ($4.5\times$ increase), while demonstrating near-linear scalability across a 10-node cluster. A case study on Renaissance manuscripts confirmed the system's ability to deliver accurate cross-modal retrieval with sub-second response times. The findings validate the framework's potential to enhance access, preservation, and research in large-scale cultural heritage archives. Future work will explore GPU-accelerated ANN search, federated retrieval across institutions, multilingual support, and 3D/XR integration to further expand its applicability and impact.

Acknowledgment

This work is supported by NFC: Institute of Engineering & Fertilizer Research, Faisalabad and Higher Education Commission (HEC) of Pakistan.

Declaration on Generative AI

The authors did not use Generative AI to create, analyze, or interpret the scientific content of this manuscript. Grammarly was used solely as a writing-assistance tool to improve grammar, spelling, clarity, and sentence structure. After using Grammarly, the authors carefully reviewed and revised the manuscript to ensure the accuracy, originality, and integrity of its content. The authors accept full responsibility for all material presented in this publication.

References

- [1] F. Ali, A. Hashem Mohammed, Content based image retrieval (cbir) by statistical methods, Baghdad Science Journal 17 (2020) 40.

- [2] M. Takami, P. Bell, B. Ommer, An approach to large scale interactive retrieval of cultural heritage, in: GCH, 2014, pp. 87–95.
- [3] B. S. J. L. Laskar, Preservation of cultural heritage in the digital age (2025).
- [4] L. Duranti, E. Shaffer, The memory of the world in the digital age: digitization and preservation, in: An International Conference on Permanent Access to Digital Documentary Heritage, UNESCO Conference Proceedings. Vancouver, 2012.
- [5] I. Parent, et al., Guidelines for the selection of digital heritage for long-term preservation: What, why, when, in: iPRES 2021, 2021.
- [6] J. Weagley, E. Gelches, J.-R. Park, Interoperability and metadata quality in digital video repositories: a study of dublin core, *Journal of Library Metadata* 10 (2010) 37–57.
- [7] A. Belhi, A. Bouras, A. K. Al-Ali, A. H. Sadka, *Data Analytics for Cultural Heritage*, Springer, 2021.
- [8] Z. Zeng, Y. Sun, W. Mao, Mccn: Multimodal coordinated clustering network for large-scale cross-modal retrieval, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 5427–5435.
- [9] Z. Ali, W. Shahzad, Epaco: a novel ant colony optimization for emerging patterns based classification, *Cluster Computing* 21 (2018) 453–467.
- [10] F. Ali, A. Hashem Mohammed, Content based image retrieval (cbir) by statistical methods, *Baghdad Science Journal* 17 (2020) 40.
- [11] S. Rani, N. Rajani, S. Reddy, Comparative study on content based image retrieval, *International Journal of Future Computer and Communication* 1 (2012) 366.
- [12] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, J. Li, Deep learning for content-based image retrieval: A comprehensive study, in: Proceedings of the 22nd ACM international conference on Multimedia, 2014, pp. 157–166.
- [13] J. Wu, Introduction to convolutional neural networks, *National Key Lab for Novel Software Technology. Nanjing University. China* 5 (2017) 495.
- [14] N. Ketkar, J. Moolayil, Convolutional neural networks, in: *Deep learning with Python: learn best practices of deep learning models with PyTorch*, Springer, 2021, pp. 197–242.
- [15] J. Chen, *Deep Learning Based Multimodal Retrieval*, Ph.D. thesis, INSA de Rennes, 2023.
- [16] Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, A survey of convolutional neural networks: analysis, applications, and prospects, *IEEE transactions on neural networks and learning systems* 33 (2021) 6999–7019.
- [17] X. Li, J. Yang, J. Ma, Recent developments of content-based image retrieval (cbir), *Neurocomputing* 452 (2021) 675–689.
- [18] X. Meng, J. Bradley, B. Yavuz, E. Sparks, S. Venkataraman, D. Liu, J. Freeman, D. Tsai, M. Amde, S. Owen, et al., Mllib: Machine learning in apache spark, *Journal of Machine Learning Research* 17 (2016) 1–7.
- [19] M. Zaharia, R. S. Xin, P. Wendell, T. Das, M. Armbrust, A. Dave, X. Meng, J. Rosen, S. Venkataraman, M. J. Franklin, et al., Apache spark: a unified engine for big data processing, *Communications of the ACM* 59 (2016) 56–65.
- [20] C. Capurro, G. Plets, J. Verheul, Digital heritage infrastructures as cultural policy instruments: Europeana and the enactment of european citizenship, *International Journal of Cultural Policy* 30 (2024) 304–324.
- [21] P. Hu, L. Zhen, D. Peng, P. Liu, Scalable deep multimodal learning for cross-modal retrieval, in: Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval, 2019, pp. 635–644.
- [22] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
- [23] M. Shafiq, Z. Gu, Deep residual learning for image recognition: A survey, *Applied sciences* 12 (2022) 8972.
- [24] S. Wang, A. Isaac, V. Charles, R. Koopman, A. Agoropoulou, T. Van Der Werf, Hierarchical structuring of cultural heritage objects within large aggregations, in: *International Conference on Theory and Practice of Digital Libraries*, Springer, 2013, pp. 247–259.

- [25] Y. WeiChen, E. WeipingWang, T. Bakker, L. L. PaulFieguth, S. L. Michael, Deep learning for instance retrieval: A survey, *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE* 45 (2023).
- [26] E. Kenny, Europeana: Cultural heritage in the digital age, in: *Migrating Heritage*, Routledge, 2016, pp. 85–94.
- [27] A. L. Silva, A. L. Terra, Cultural heritage on the semantic web: the europeana data model, *IFLA journal* 50 (2024) 93–107.
- [28] I. U. Rehman, Z. Ali, Z. Jan, M. Rashid, A. Abbas, N. Tariq, Deep learning empowered classification of augmented cultural heritage images., in: *VIPERC*, 2023.
- [29] P. Turan, C. Sganga, et al., Common european data space for cultural heritage: Is the eu taking the “high road” from “a single access point” to “a single market for data” for digital cultural content?, *JOURNAL OF INTELLECTUAL PROPERTY, INFORMATION TECHNOLOGY AND ELECTRONIC COMMERCE LAW* 16 (2025) 133–153.
- [30] I. U. Rehman, Z. Ali, Z. Jan, M. Rashid, W. Ali, Empirical performance analysis of classification methods on cultural heritage database., in: *VIPERC*, 2022.
- [31] M. Armbrust, T. Das, J. Torres, B. Yavuz, S. Zhu, R. Xin, A. Ghodsi, I. Stoica, M. Zaharia, Structured streaming: A declarative api for real-time applications in apache spark, in: *Proceedings of the 2018 International Conference on Management of Data*, 2018, pp. 601–613.
- [32] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, in: *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies*, volume 1 (long and short papers), 2019, pp. 4171–4186.
- [33] Y. A. Malkov, D. A. Yashunin, Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs, *IEEE transactions on pattern analysis and machine intelligence* 42 (2018) 824–836.
- [34] B. Dixit, R. Kuc, M. Rogozinski, S. Chhajed, *Elasticsearch: A Complete Guide*, Packt Publishing Ltd, 2017.
- [35] N. Freire, R. Voorburg, R. Cornelissen, S. De Valk, E. Meijers, A. Isaac, Aggregation of linked data in the cultural heritage domain: A case study in the europeana network, *Information* 10 (2019) 252.
- [36] A. Isaac, B. Haslhofer, Europeana linked open data–data. europeana. eu, *Semantic Web* 4 (2013) 291–297.
- [37] B. Haslhofer, A. Isaac, data. europeana. eu: The europeana linked open data pilot, in: *Proceedings of the international conference on Dublin Core and metadata applications*, Dublin Core Metadata Initiative, 2011.
- [38] Z. Zeng, Y. Sun, W. Mao, Mccn: Multimodal coordinated clustering network for large-scale cross-modal retrieval, in: *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 5427–5435.